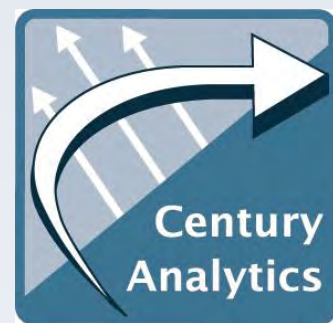# Impacts of Study Island on Student Mathematics Achievement

June 2019

Bruce Randel, Ph.D.
Century Analytics, Inc.
Centennial CO
www.centuryanalytics.com

Executive Summary

Century Analytics conducted a rigorous evaluation of Edmentum's Study Island to estimate its impact on student achievement in mathematics in Grade 2 through Grade 8. Study Island is a flexible product that provides students with practice and instructional support on standards-based topics in mathematics, reading, science, and social studies. This study's quasi-experimental design (QED), analyses, and measures meet the What Works Clearinghouse (WWC) 4.0 standards needed to achieve a rating of Meets WWC Group Design Standards with Reservations (WWC, 2017). This study also meets the Every Student Succeeds Act (ESSA) guidance for Moderate Evidence (U.S. Department of Education, 2016).

Students were included in the Study Island group if they had completed at least half of the Study Island practice topics available in their grade level between their fall and winter administration of the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) assessment. The MAP is a standardized educational assessment that meets WWC standards for validity and reliability. Students who had completed no Study Island topics during the same period were matched to the Study Island students in each grade level on their fall MAP test scores and the duration between their fall and winter MAP test administrations. At each grade level, Study Island students were within the WWC threshold for baseline equivalence on both fall MAP test scores and duration.

Analyses revealed statistically significant positive impacts of Study Island use on winter MAP math achievement in Grades 2 through 7. Effect sizes ranged from 0.26 for Grade 2 to .07 for Grade 7 and improvement indexes ranging from 10.26 to 2.79. Improvement indexes show the expected change in percentile rank for an average comparison student if he or she had been in the intervention group.

Results suggest that students who complete at least half of the Study Island topics available in their grade level will make statistically significant gains in achievement relative to students who do not complete any Study Island topics. These results suggest that Study Island is providing students with practice in standards-based academic content that improves students' scores on standardized tests.

This study is not without limitations. The definition for the Study Island group focused solely on topic completion. The study does not shed any light on the potential impact of any other of Study Island's student resources (e.g., lessons, games, flash cards, printable worksheets) or the impact of Study Island when integrated into classroom instruction.

Many students in the Study Island group completed more than the minimum number of topics needed to be included in the Study Island group. In some cases, students completed more than double the number of topics available at their grade level. Whether this amount of practice can be attributed to Study Island or some characteristics of the students is a question unanswered by this study. In addition, the lack of any student demographic variables limits the generalizability of the study's findings.

Future research on Study Island should incorporate a broader definition of student usage in order to estimate the impact of the many student resources available beyond practice topics. This research also should include student demographic characteristics to help understand which groups of students may benefit most from Study Island and to support generalizing study findings. Although Study Island is designed primarily to provide students with practice on standards-based academic topics, it also has resources for teachers to help integrate Study Island into classroom instruction. Future research also should include student-level criteria and classroom-level criteria for inclusion in the Study Island group to better understand the impacts of Study Island on student achievement.

Introduction

Study Island is a flexible formative assessment tool that can be used for practice and instructional support on standards-based topics, test preparation, classroom assessment, and intervention support in four academic subjects: math, reading, science, and social studies (Edmentum, 2018). Study Island provides students with in-depth practice, immediate feedback on progress, and remediation when needed.

Study Island is structured around academic topics: a grouping of conceptual material within a subject area and grade aligned to one or more state standards. For each subject and grade, the topics cover the state standards in their entirety. Each Study Island topic includes practice items, lessons, and supporting instructional materials (e.g., games, flash cards, printable worksheets). If a student works through all topics in a grade and subject area, he/she will have practiced the full range of learning goals covered by the standards.

Students gain practice on topics by completing practice sessions. Practice sessions are online formative assessments composed of at least 10 multiple-choice practice items or technology enhanced practice items that are scored online. Students can earn virtual trophies by completing a practice session with a score of 70% correct or more. Students receive immediate feedback on their answer to each item in the practice session. Explanations of the correct answer appear if the student gets the item wrong, providing real-time remediation based on individual student performance. Students can revisit any items they answer incorrectly at the end of each practice session.

Study Island's real-time feedback and links to supportive instructional materials provide students with remediation based on student performance. For example, if the student scores 40% or lower on a practice activity, Study Island routes the student to practice sessions for lower level topics that are building blocks for higher-level topics. Once the student demonstrates success at the lower level, they are routed back up to the higher level.

Study Island allows students to learn at their own pace, choosing which topics to work on in addition to topics assigned to them by their teachers. The program provides students with access to a variety of supporting instructional materials targeted to their learning needs. In this way, Study Island provides each student with an individual learning trajectory.

In addition to practice items and instructional materials targeted at individual students' needs, Study Island also provides a variety of features to support teachers and administrators. Teachers can customize Study Island practice to individual students by assigning students specific topics, selecting specific practice items, and customizing the score needed to earn virtual trophies. Teachers also can access lessons and supporting instructional materials aligned to each topic at the beginning and/or end of the practice activity.

Study Island's array of customizable reports allow teachers and administrators to identify student weaknesses and monitor student progress toward the mastery of standards-based content. If students fall behind or advance ahead, teachers can assign additional practice targeted to individual students while continuing to monitor progress. Teachers also can use information from Study Island to inform instruction or remediation within the classroom. With the formative information from Study Island integrated into classroom instruction, teachers can create individualized instructional assignments based on demonstrated student need to help students meet specific content benchmarks and standards.

One of the supporting features available to Study Island customers is Study Island's NWEA™ MAP® Link. This feature allows customers to upload student MAP scores and integrate them with Study Island to help target specific Study Island content and automatically target learner's individual needs. The program is available for Grade 2 through high school.

The following research question guided the design and analyses used in this study:

>What is the impact of Study Island usage on student mathematics achievement in Grade 2 through Grade 8 relative to students who did not use Study Island?

Method

The purpose of this study was to provide a rigorous estimate of the impact of Study Island use on student achievement in mathematics. Rigorous studies of educational interventions and estimates of impacts are needed by state and local education agencies to select and implement interventions that improve academic outcomes for students (U.S. Department of Education, 2016).

The study was designed to meet the What Works Clearinghouse (WWC) 4.0 standards for quasi-experimental designs (QED) necessary to achieve a rating of *Meets WWC Group Design Standards with Reservations* (WWC, 2017). In meeting WWC standards, the study also was designed to meet the requirements of the Every Student Succeeds Act (ESSA) guidance for *Moderate Evidence* (U.S. Department of Education, 2016).

Data

Century Analytics obtained student data from Edmentum to conduct this study. These data included unique student identifiers, student grade level, identifiers for academic subjects, Study Island usage data, number of Study Island topics completed, and scores from the fall and winter administrations of the NWEA MAP assessment. No student demographic variables other than location were available for analysis. The study used data on fall MAP scores as the baseline measure and winter MAP scores as the outcome measure.

Study Island usage data. Data were compiled to summarize the number of distinct Study Island topics each student completed between the baseline (fall) MAP assessment and the outcome (winter) MAP assessment. Topics completed below and above students' grade level were included to capture personalized remedial or enrichment practice using Study Island. Actual assessments of topics (e.g., diagnostic tests) were not included in the number of topics completed. Practice sessions that were not completed also were not included. Customer-generated topics or customer-generated practice questions, although very rare, were filtered out. Any topics where the student was practicing in Study Island's "Game Mode," in which the student is motivated to complete practice topics with brief, interstitial video games, were excluded to ensure that the total number of topics completed was based only on topic practice that was directed towards improving academic skills.

NWEA MAP data. The data provided also included NWEA MAP scores uploaded to Study Island by Study Island customers. Data cleaning rules were applied to ensure the quality of the test score data. Only students with a valid test score of 100 or greater were included. Student records were dropped if multiple MAP scores were uploaded for a given content area and testing period in order to identify one

score per test and student, per season. The MAP score data were merged to the Study Island usage data by student ID.

Measures

Student achievement, both at baseline (fall) and outcome (winter) was measured using the NWEA MAP. The MAP is a computer adaptive interim assessment that measures achievement in mathematics, reading, and other subjects for students in Grades 2 through 12 (NWEA, 2011). The MAP is designed to measure growth over time with all items anchored to a vertically-aligned equal-interval scale (RIT scale) across grade levels to provide a longitudinal measure of student growth. As such, the MAP provides scores for student achievement even if the student is performing below or above typical grade level expectations. The MAP is aligned to state standards and covers a range of skills across grade levels. The MAP meets the WWC standards for outcomes in terms of validity and reliability. Because the MAP measures content aligned to national and state standards it is not over-aligned to the Study Island intervention.

Design

This study used a quasi-experimental design (QED) in order to meet WWC (4.0) standards with reservations. According to the WWC, a QED uses a non-random process to form the intervention and comparison conditions (WWC, 2017). The WWC allows groups to be formed using a variety of methods as long as the groups are mutually exclusive. That is, units (e.g., students or schools) can only be analyzed as a member of a single group. Further, in a QED, the WWC accepts assignment to the intervention based on observed characteristics. Assignment to study conditions for this study was conducted at the student level.

Students included in this study were selected from Grade 2 through Grade 8 students enrolled in a Study Island MAP Link program during the 2018-19 school year. The sample excludes students enrolled in the Study Island Benchmark program. To be eligible for inclusion in the study, student records had to meet the following eligibility criteria:

- Baseline MAP scores occurring between August 5, 2018 and September 16, 2018;
- Outcome MAP scores occurring between November 25, 2018 and January 20, 2019; and
- Duration between baseline and outcome test score dates of 84 to 140 days.

This study focused on students in Grades 2 through 8. Students in Kindergarten and Grade 1 were not included in the study because the Study Island MAP Link program does not include mappings between NWEA scores and Study Island topics for kindergarten and first grade. Grades 9 and above were not included in the study because sufficient sample sizes were not available.

Students meeting the eligibility criteria were then selected for inclusion in either the Study Island (intervention) group or the comparison group. The Study Island group included students who completed a total number of distinct topics equal to at least half of the available topics for that grade (see Table 1). The rationale for this is that students who have completed half of the available topics approximately halfway through the school year are on track to have completed all available topics by the end of the year and therefore have been exposed to a broad range of topics that cover grade level expectations. Topics from below the student's grade level (remediation) or above the student's grade level (enrichment) were included as contributing to the total number completed because these represent

practice targeted to students' individual skill levels. (See Appendix A for the number of lessons completed by students in the Study Island group.)

Table 1. Study Island (Treatment) Group Criterion

Grade	Minimum Math Topics Completed
2	11
3	15
4	18
5	14
6	18
7	13
8	13

Students eligible for inclusion in the comparison group included those students who met the above eligibility criteria and had completed zero Study Island topics. Propensity score matching was used to match eligible comparison students to Study Island group students separately by grade. One-to-four matching—except in Grade 3, which used one-to-three matching—was conducted using the nearest neighbor method and a logistic regression model with the fall MAP score and duration (time between fall and winter MAP test administrations) as the matching variables. Duration is an exogenous variable (i.e., not related to or affected by group selection) because it was based on the MAP test score data uploaded to Study Island by customers prior to the conduct of this study.

Students in both the Study Island group and the comparison group were administered the MAP and were exposed to any features or services available through the MAP. Thus, the difference between the two groups in terms of the educational instruction and implementation (i.e., the intervention contrast) was use of Study Island. In this way, this study design estimates the impact of student use of Study Island, defined as completion of the prescribed number of topics.

Baseline Equivalence

In order to meet WWC standards with reservation for a QED, baseline equivalence must be established for the analytic samples of the Study Island (intervention) and comparison groups. In addition, baseline equivalence needs to be established separately for each grade level included in the analyses. Finally, baseline equivalence must be established using a measure that meets WWC standards.

Baseline equivalence was established using fall MAP scores and duration separately for the analytics sample in each grade level. As described above in the Measures section, the MAP assessment meets WWC standards for baseline and outcome measures. All grades had baseline differences between students in the Study Island and comparison groups that were under the WWC threshold for baseline equivalence (i.e., ≤ 0.25 standard deviation) using a WWC accepted method for calculating baseline differences (WWC, 2017). Appendix B provides the statistics for baseline equivalence. Also included in Appendix B is information on the distribution of students in both groups across states.

Analyses and Results

Data were analyzed to estimate differences between students in the Study Island (intervention) group and the comparison group on the winter MAP mathematics assessment. Impact analyses were conducted using the following linear regression model fit to the data separately for each grade level.

$$Y_i = \beta_0 + \beta_1(\text{GROUP})_i + \beta_2(\text{FALLMAP})_i + \beta_3(\text{DURATION})_i + e_i$$

Where: Y_i is student i's winter math MAP score. β_0 is the regression adjusted comparison group mean. β_1 is the adjusted mean difference between the intervention and comparison groups, and GROUP represents the group status of student i coded as 0 = comparison and 1 = intervention. β_2 is the regression slope for the fall (baseline) MAP math score and β_3 is the regression slope for duration. FALLMAP is student i's fall MAP score in math, and DURATION is the number of days between the fall MAP administration and winter MAP administration for student i. e_i is the residual for student i.

Impact analyses yielded statistically significant positive impacts for all grade levels in mathematics except Grade 8 (Table 2). Output from the regression analyses are provided in Appendix C. Adjusted mean differences between the intervention and comparison groups ranged from 0.52 points on the MAP RIT scale for Grade 8 to 3.28 points for Grade 2. These differences translate into effect sizes ranging from 0.03 for Grade 8 to 0.26 for Grade 2.

Table 2. Impacts on Mathematics Achievement

	N	Mean MAP Score	SD	Adjusted Mean Difference (SE)	Pooled Standard Deviation	Effect Size	Improve Index
Grade 2							
Comparison	1340	187.62	12.55	3.28***	12.49	0.26	10.26
Study Island	335	190.90	12.23	(0.40)			
Grade 3							
Comparison	2919	197.33	10.95	2.66***	10.94	0.24	9.48
Study Island	973	199.99	10.93	(0.21)			
Grade 4							
Comparison	3208	206.20	11.87	1.98***	11.81	0.17	6.75
Study Island	802	208.18	11.58	(0.23)			
Grade 5							
Comparison	3684	213.72	14.38	1.67***	14.51	0.12	4.78
Study Island	921	215.39	15.03	(0.23)			
Grade 6							
Comparison	2360	214.40	15.41	1.55***	15.56	0.10	3.98
Study Island	590	215.94	16.14	(0.29)			
Grade 7							
Comparison	1520	220.55	16.82	1.10**	16.78	0.07	2.79
Study Island	380	221.65	16.60	(0.40)			
Grade 8							
Comparison	1560	227.16	19.41	0.52	19.65	0.03	1.20
Study Island	390	227.68	20.56	(0.42)			

Note. MAP scores are on the RIT scale. Numbers in table rounded to 2 decimals. M = mean. SD = standard deviation. SE = Standard error. Improve Index = Improvement index. Effect size measured as Hedges' g.
** = p-value < .01. *** = p-value < .001.

In addition to translating the impacts of Study Island use into effect sizes, the improvement index is another useful method to aid in the interpretation of the practical importance of impacts. The improvement index represents the difference in percentile rank at the mean (i.e., the 50th percentile) between the intervention group and the comparison group (WWC, 2017). The improvement index shows the expected change in percentile rank for an average comparison student if he or she had received the intervention. Percentile improvements for Study Island use in math ranged from 1.20 for Grade 8 to 10.26 for Grade 2. This largest improvement index is equivalent to a comparison student improving from the 50th percentile to the 60th percentile.

Finally, normative data from the MAP can help interpret results from this study. These normative data show the average achievement levels by grade at the beginning, middle, and end of the school year based on data collected from 2011 to 2014 (NWEA, 2015). The MAP norms also provide educators with information on the normative growth on the MAP assessment.

The average achievement levels from the norms can be compared to the average achievement for the study groups on the winter administration of the MAP (Table 3). These data show that the students in this study had similar levels of achievement to the national norms. In the lower grades, students participating in the study had slightly higher average scores than the norms, and in the upper grades study students had average achievement slightly lower than the norms. Rather than indicating impacts, the comparison of achievement levels between the students in this study and the norms shows that the students in this study are likely to represent typical students in terms of achievement and not either especially low or especially high achieving students. This suggests this study's results are broadly generalizable.

Table 3. Average Achievement from NWEA Norms and Study Island Groups by Grade

Grade	Middle Year MAP Norm	Comparison Group Average Winter Score	Study Island Group Average Winter Score	Study Island Impacts
2	186.4	187.62	190.90	3.28***
3	198.2	197.33	199.99	2.66***
4	208.7	206.20	208.18	1.98***
5	217.2	213.72	215.39	1.67***
6	222.1	214.40	215.94	1.55***
7	226.1	220.55	221.65	1.10**
8	229.1	227.16	227.68	0.52

Notes. All scores are on the RIT scale. Normative data from NWEA (2015).
** = p-value < .01. *** = p-value < .001.

In addition to the average achievement provided by the norms, data on normative growth can be used to aid in the interpretation of the impacts found in this study. NWEA calculated the average amount of growth students show on the MAP between the beginning of the year and middle of the year administrations (the same time period used in this study). This normative growth is shown in the first column in Table 4. These normative growth values can be compared to the impacts of Study Island to provide some additional meaning to the difference in scores between students who used Study Island and those in the comparison group. Dividing the impact by the normative growth shows the impacts in terms of a percentage of normative growth (last column of Table 4). Using this method, the impacts of Study Island represent approximately 30% of the typical growth seen by students during the first half of the school year, except for Grade 8 where no statically significant difference was found. In grade 2 for

example, students in the Study Island group scored 3.28 points higher than students in the comparison group. This 3.28 point difference represents 35% of the typical growth on the MAP for Grade 2 students.

Table 4. MAP Normative Growth and Study Island Impacts by Grade

Grade	Beginning to Middle Year Normative Growth	Study Island Impacts	Impact as Percent of Normative Growth
2	9.5	3.28***	35%
3	7.8	2.66***	34%
4	6.8	1.98***	29%
5	5.8	1.67***	29%
6	4.4	1.55***	35%
7	3.5	1.10**	31%
8	2.9	0.52	18%

Notes. All scores are on the RIT scale. Normative data from NWEA (2015).
** = p-value < .01. *** = p-value < .001.

Interpretation of the Study Island impacts as a percentage of normative growth should not be confused with a difference in growth between Study Island and comparison group students. For example, Study Island students in 2nd grade did **not** grow 35% more than comparison group students. Study Island and comparison group students were not compared in terms of growth but only compared on their average winter MAP assessment scores (regression adjusted).

<div align="center">

Summary

</div>

Using a rigorous design and analytic strategy, this study estimated the impact of Study Island use on students' mathematics achievement as measured by the Measures of Academic Progress (MAP) at the level of rigor needed to meet WWC standards with reservations (WWC, 2017). Baseline equivalence was established between the Study Island (intervention) group and the comparison group. The Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) was used as the baseline measure and the outcome measure of student mathematics achievement. The MAP is a standardized educational assessment that meets WWC standards for validity and reliability. The baseline and outcome measures are aligned to national and state academic content standards and so are not over-aligned to the Study Island intervention. The study had no confounds.

The study also meets criteria set forth by the Every Students Succeeds Act (U.S. Department of Education, 2016). The Department of Education considers a quasi-experimental study to be "well-designed and well-implemented" if it receives a *Meets WWC Design Standards with Reservations* rating or is of equal quality (U.S. Department of Education, 2016). The study also meets the ESSA criteria for statistically significant positive effects. These two aspects of the study mean it qualifies as providing Moderate evidence (Level 2) of Study Island's effectiveness.

Study Island had a statistically significant impact on student mathematics achievement at every grade level analyzed except Grade 8. These impacts occurred on the winter administrations of the MAP assessment. Students who met the definition for inclusion in the Study Island (intervention) group—completion of at least half the grade level topics in math between their fall and winter MAP assessments—showed greater gains in mathematics achievement than students who completed zero Study Island topics.

The results of this study suggest that students who use Study Island and complete topics in math will make gains in achievement relative to students who do not complete any topics. The statistically significant gains made by students in the Study Island group over those students in the comparison group also suggest that practice on academic topics using Study Island helps students improve their mathematics knowledge and skills.

The data on topic completion in Appendix A show that many students completed many more than the minimum number of topics needed to be included in the Study Island group. In some cases, students completed more than double the number of topics available at their grade level. These data suggest that many of the students in the Study Island group spent a considerable amount of time practicing academic content, perhaps more than can be reasonably expected of typical students. Few would doubt that time spent practicing academic content, in an online testing format, is positively correlated with test scores. The question unanswered by this study is whether Study Island is the cause of this time spent in practice, if the practice time is driven by student characteristics, or some combination of the two.

Limitations

This study has several limitations worth noting. It used a focused definition for the intervention group: students who had completed at least half of the available Study Island topics for a specific grade (Table 1). But Study Island involves much more than practice of academic topics, and use of Study Island cannot simply be measured as completion of topics. Study Island includes additional resources, such as lessons, games, flash cards, and printable worksheets; summarizing Study Island use as simply the number of 10-item practice sessions completed is likely providing an incomplete picture of usage.

This study did not estimate the impact of use of any of Study Island's additional resources. Students in the Study Island group, or in the comparison group, may or may not have used these additional resources to varying degrees. The use, or lack of use, of these resources could be influencing student achievement and, therefore, the findings in this study. Given the study design, it is possible that the key difference between the Study Island group students and the comparison group students was not completion of topics but use of additional resources, and the additional resources are, in fact, responsible for the higher levels of achievement among the Study Island students.

This study used a design sufficient to meet WWC standards with reservations. The Study Island students and comparison students were equivalent at baseline (fall MAP scores) on mathematics achievement. Students in the two groups were also equivalent in terms of the duration between their fall and winter MAP administrations. Both the students' fall math scores and their test durations were used as statistical adjustments for estimating impacts on math. No other student characteristics, however, were available or included in the study.

This study used a rigorous quasi-experimental design (QED) that is acceptable to meet WWC standards with reservations. Along with the statistically significant positive impacts, this study meets ESSA Level 2 standards. That said, the study was unable to control for student characteristics other than baseline (fall) achievement. It is possible that students in the two groups differed on some key characteristic(s), and this difference was responsible for the difference in winter MAP scores rather than the difference in completion of Study Island topics. The lack of student demographic characteristics also limits the generalizability of the study results. It is unclear from this study what types of student were included in the Study Island group or if students of differing backgrounds experienced differing impacts from Study Island usage.

Students were the unit of assignment to the intervention and comparison groups. Study Island usage typically differs by students, so using students as the unit of assignment is appropriate. Study Island, however, is not used by students in isolation. Study Island also makes resources available to teachers and can be used to supplement regular classroom instruction. Teachers can use Study Island to assign individual students practice in areas of need, select specific practice items, and revise the score needed to advance to the next topic. Like the additional resources available to individual students, the features of Study Island for classroom and school use likely have an effect on student achievement. This study, however, was unable to estimate the impacts of any of these additional features of Study Island.

Further Research

This study provides a rigorous estimate of the impact of student completion of Study Island topics on student achievement in mathematics. Additional research is needed to understand how other features of Study Island impact student achievement. Future research also should address the limitations of this study. In addition to including student demographic characteristics as part of future analyses, further research also should examine other aspects of student usage and how these might impact student achievement. These could include student use of lessons, flash cards, and printable worksheets.

Although Study Island is designed primarily as a resource for students to gain practice and instructional support on standards-based topics as a means of test preparation, Study Island is still used within the classroom context. Integration of students' use of Study Island into classroom instruction likely magnifies the benefits of its use. A future study at the classroom level, that includes both student-level criteria and classroom-level criteria for inclusion in the Study Island group, could shed additional light on the impacts of Study Island on student achievement.

An unbiased estimate of Study Island's impact can only be provided by a randomized controlled trial (RCT). In this type of study, students (or classrooms) are randomly assigned to either use Study Island or conduct business as usual, creating two groups that are equivalent in expectation on all characteristics, known and unknown. This equivalence means any difference in achievement between the study groups can be attributed to Study Island usage. A well-conducted RCT eliminates the possibility that differences between intervention and comparison groups on outcome measures are caused by differences in baseline characteristics rather than the intervention under study, a limitation of the present study.

Finally, impact studies are best conducted in parallel with studies of implementation fidelity. Findings from the two studies complement each other and aid in the interpretation of results. Studies of implementation fidelity inform the impact research by aiding in the definition of intervention group and communicating to the research audience what level of usage resulted in the impacts. Studies of impact inform implementation research by estimating impacts at different levels of implementation and helping to focus on how much usage is needed to produce statistically significant and meaningful increases in student achievement.

References

Edmentum. (2018). *Research report: Examination of Study Island practice and benchmarks for Allentown School District*. Bloomington, MN: Author.

Northwest Evaluation Association. (2011). *Technical manual for Measures of Academic Progress® (MAP®) and Measures of Academic Progress for Primary Grades (MPG)*. Portland, OR: Author.

Northwest Evaluation Association. (2015). *2015 NWEA MAP Growth normative data*. Retrieved from https://www.nwea.org/resource-library/research/2015-normative-data-3

U.S. Department of Education, Office of Elementary and Secondary Education. (2016, September). *Non-regulatory guidance: Using evidence to strengthen education investments*. Retrieved from https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse: Standards handbook (Version 4.0).* Retrieved from http://whatworks.ed.gov

Table A.1. Number of Study Island Students Completing Mathematics Topics by Grade Level

Topics completed	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
11 Topics	49						
12 Topics	49						
13 Topics	37					41	42
14 Topics	30			156		48	57
15 Topics	27	118		125		33	45
16 Topics	18	108		86		43	27
17 Topics	21	84		90		20	34
18 Topics	18	67	95	67	69	37	26
19 Topics	19	65	68	51	50	16	20
20 Topics	14	70	65	49	48	21	15
21 Topics	6	55	46	38	38	19	15
22 Topics	9	54	49	35	37	15	14
23 Topics	6	32	39	23	37	7	12
24 Topics	5	32	48	28	33	15	12
25 Topics	0	31	44	19	31	6	8
26 Topics	4	22	35	17	22	8	2
27 Topics	1	18	40	16	26	5	8
28 Topics	3	23	28	12	17	5	6
29 Topics	1	17	29	15	22	7	7
30 Topics	0	20	23	10	9	4	5
31-40 Topics	12	95	148	56	92	26	21
41-50 Topics	2	36	23	20	42	1	10
51 or more Topics	4	26	22	8	17	3	4
Total	335	973	802	921	590	380	390

Note. Minimum number of topics completed varies by grade level. See Table 1.

Appendix B
Baseline Equivalence

Table B.1. Fall MAP Mathematics Baseline Equivalence by Grade

	N	M	SD	Adjusted Mean Difference	Pooled SD	Effect Size
Grade 2						
Comparison	1340	179.28	13.32	0.40	13.40	0.03
Study Island	335	179.69	13.72			
Grade 3						
Comparison	2919	190.34	10.96	0.71	11.09	0.06
Study Island	973	191.05	11.48			
Grade 4						
Comparison	3208	201.00	11.81	0.62	11.85	0.05
Study Island	802	201.62	12.02			
Grade 5						
Comparison	3684	209.01	13.90	0.80	14.14	0.06
Study Island	921	209.81	15.08			
Grade 6						
Comparison	2360	210.29	14.96	2.75	15.01	0.18
Study Island	590	213.03	15.23			
Grade 7						
Comparison	1520	216.75	17.02	0.68	17.06	0.04
Study Island	380	217.43	17.22			
Grade 8						
Comparison	1560	224.24	19.30	-0.18	19.37	-0.01
Study Island	390	224.06	19.68			

Note. Numbers in table rounded to 2 decimals. M = mean. SD = standard deviation. Effect size measured as Hedges' *g*.

Table B.2. Duration Baseline Equivalence by Grade

	N	M	SD	Adjusted Mean Difference	Pooled SD	Effect Size
Grade 2						
Comparison	1340	112.69	13.72	-0.64	13.34	-0.05
Study Island	335	112.04	11.68			
Grade 3						
Comparison	2919	109.18	14.11	0.25	13.71	0.02
Study Island	973	109.43	12.42			
Grade 4						
Comparison	3208	106.81	13.21	0.38	13.03	0.03
Study Island	802	107.19	12.29			
Grade 5						
Comparison	3684	111.18	14.14	0.26	13.96	0.02
Study Island	921	111.44	13.19			
Grade 6						
Comparison	2360	113.29	14.55	-1.01	14.44	-0.07
Study Island	590	112.28	13.99			
Grade 7						
Comparison	1520	111.06	13.25	0.20	13.17	0.02
Study Island	380	111.26	12.81			
Grade 8						
Comparison	1560	110.09	13.00	-0.08	13.04	-0.01
Study Island	390	110.01	13.19			

Note. Numbers in table rounded to 2 decimals. M = mean. SD = standard deviation. Effect size measured as Hedges' *g*.

Table B.3. Student Island and Comparison Group Students by Grade and State

State	Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	SI	C	SI	C	SI	C	SI	C	SI	C	SI	C	SI	C
AR			29	6	20	7	55	1	19	1	12		3	
AZ			7	92	1	195	7	95	5	98	1	38	5	60
CA		1	48	9	50	9	89	10	67	11	35	6	80	11
FL				1										
GA	161	52	134	176	160	118	230	179	298	97	356	57	314	26
IL	13	23	50	57	74	35	28	22	25	19	14	1	17	8
IN	349	91	675	167	493	87	336	120	239	142	253	64	360	106
KS	38		132		116	16	208	66	56		61		61	
KY	119	23	379	163	531	143	651	57	392	52	254	93	318	72
MI		2	2		1		13	8	76	3	17	19		
MN							1							
MO	2	6			1		28	4	4	8	1	17	1	53
MS	4	11	10	31	6	14	22	20	42	36	4	15	5	11
NE	7		40		51		70	3	54		17	1	19	
OH	469	107	1,108	185	1,207	131	1,411	184	954	38	413	50	315	7
OK									2		18		29	
PA	35		52	21	29	7	47	30	16	2	9		1	1
SC	30	5	62	21	120	28		8						
SD	6		10		6		6		2	1	2			
TN							14	1	1		2		4	
TX	107	14	181	44	342	12	409	60	80	32	51	19	28	35
WI							59	53	28	50				
Total	1,340	335	2,919	973	3,208	802	3,684	921	2,360	590	1,520	380	1,560	390

Note. SI = Study Island group. C = comparison group.

Table C.1. Grade 2 Estimates of Regression Coefficients

Parameter	Coefficient	Std. Error	t value	p-value	95% Conf. Int.	
Study Island	3.28	0.40	8.18	0.00	2.49	4.07
Fall MAP score	0.79	0.01	66.18	0.00	0.77	0.82
Duration	0.03	0.01	2.15	0.03	0.00	0.05
Intercept	42.63	2.53	16.83	0.00	37.66	47.60

Std. Error = standard error
95% Conf. Int. = 95% confidence interval

Table C.1. Grade 3 Estimates of Regression Coefficients

Parameter	Coefficient	Std. Error	t value	p-value	95% Conf. Int.	
Study Island	2.66	0.21	12.58	0.00	2.25	3.07
Fall MAP score	0.84	0.01	101.98	0.00	0.83	0.86
Duration	0.04	0.01	6.67	0.00	0.03	0.06
Intercept	32.25	1.76	18.33	0.00	28.80	35.70

Std. Error = standard error
95% Conf. Int. = 95% confidence interval

Table C.1. Grade 4 Estimates of Regression Coefficients

Parameter	Coefficient	Std. Error	t value	p-value	95% Conf. Int.	
Study Island	1.98	0.23	8.5	0.00	1.52	2.43
Fall MAP score	0.86	0.01	108.3	0.00	0.84	0.87
Duration	0.05	0.01	7.47	0.00	0.04	0.07
Intercept	28.53	1.68	16.95	0.00	25.23	31.83

Std. Error = standard error
95% Conf. Int. = 95% confidence interval

Table C.1. Grade 5 Estimates of Regression Coefficients

Parameter	Coefficient	Std. Error	t value	p-value	95% Conf. Int.	
Study Island	1.67	0.23	7.3	0.00	1.22	2.12
Fall MAP score	0.93	0.01	142.96	0.00	0.91	0.94
Duration	0.04	0.01	6.05	0.00	0.03	0.05
Intercept	15.57	1.55	10.02	0.00	12.52	18.62

Std. Error = standard error
95% Conf. Int. = 95% confidence interval

Table C.1. Grade 6 Estimates of Regression Coefficients

Parameter	Coefficient	Std. Error	t value	p-value	95% Conf. Int.	
Study Island	1.55	0.29	5.25	0.00	0.97	2.12
Fall MAP score	0.94	0.01	120.77	0.00	0.93	0.96
Duration	0.02	0.01	2.59	0.01	0.01	0.04
Intercept	13.29	1.86	7.13	0.00	9.63	16.94

Std. Error = standard error
95% Conf. Int. = 95% confidence interval

Table C.1. Grade 7 Estimates of Regression Coefficients

Parameter	Coefficient	Std. Error	*t* value	*p*-value	95% Conf. Int.	
Study Island	1.10	0.40	2.72	0.01	0.31	1.89
Fall MAP score	0.89	0.01	92.02	0.00	0.87	0.91
Duration	0.01	0.01	1.12	0.26	-0.01	0.04
Intercept	25.92	2.26	11.47	0.00	21.49	30.36

Std. Error = standard error
95% Conf. Int. = 95% confidence interval

Table C.1. Grade 8 Estimates of Regression Coefficients

Parameter	Coefficient	Std. Error	*t* value	*p*-value	95% Conf. Int.	
Study Island	0.52	0.42	1.24	0.22	-0.30	1.34
Fall MAP score	0.94	0.01	105.65	0.00	0.92	0.95
Duration	0.02	0.01	1.55	0.12	-0.01	0.05
Intercept	14.90	2.20	6.78	0.00	10.59	19.21

Std. Error = standard error
95% Conf. Int. = 95% confidence interval